DOCUMENT RESUME

ED 150 185 TM 006 914

AUTHOR Hardy, Roy
TITLE From CIRCUS to EL CIRCO: Issues in Instrument Development for Young Spanish-speaking Children.
PUB DATE Feb 76
NOTE 19p.; Paper presented at the National Conference on Multicultural Curriculum and Materials (3rd, San Francisco, California, February 19-21, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Cubans; *Culture Free Tests; *Diagnostic Tests; Early Childhood Education; Mexican Americans; Norms; *Preschool Tests; Puerto Ricans; *Spanish Speaking; Standardized Tests; Test Bias; *Test Construction; *Testing Problems; Test Interpretation
IDENTIFIERS CIRCO Assessment Battery; Circus Assessment Battery

ABSTRACT

The CIRCO project is a large scale effort to design a series of diagnostic instruments, based on the CIRCUS tests, for Spanish-speaking children in preschool, kindergarten, and first grade classrooms in the United States. The goal is to develop measures with the following characteristics: (1) is suitable for use with Spanish-speaking children from different cultural and language dialect backgrounds; (2) emphasizes a differentiated view of human abilities; (3) eliminates cultural or regional biases, with sufficient pilot testing; (4) has norms representative of the Spanish-speaking population of 4-6 year-old children in preschool programs; (5) considers the construct of Spanish as a language, and the syntax and lexical differences between Spanish and English; (6) adequately guards against test score misinterpretation and test misuse. In both the CIRCUS and CIRCO systems, interpretive scores (such as grade equivalent scores) will be abandoned, and test publishers will provide more direct interpretation with the score report. The plans for pilot testing, field testing, norming, and item review are described. (Author/MV)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED150185

From CIRCUS TO EL CIRCO: Issues in Instrument Development for Young Spanish-speaking Children

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Roy Hardy

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Roy Hardy
Educational Testing Service
Southern Office
Atlanta, Georgia

a paper presented to

The Third National Conference on
Multicultural Curriculum and Materials

February 19-121, 1976
San Francisco, California

TM006 914

From CIRCUS TO EL CIRCO: Issues in Instrument Development for Young Spanish-speaking Children

Roy Hardy
Educational Testing Service

The CIRCO project is a large scale effort to develop a series of assessment instruments for Spanish-speaking children in preschool, kindergarten, and first grade classrooms in the United States. Over a three-year period we will administer over 45,000 tests to more than 15,000 Spanish-speaking children as a part of the development process. The final battery will include at least 13 different student measures. As Dr. Bernal has pointed out, the CIRCO project is in response to a critical and intense need. Although the need is immediate, the development of tests with the properties of reliability, validity, and relevance as described by Dr. Anderson is not an overnight process.

Our commitment is to the development of assessment methods that will avoid many of the faults and weaknesses of presently available measures and that will meet many of the standards recommended by bilingual educators for test developments for Spanish-speaking children. The goal of our project is to develop measures with the following characteristics:

1. Suitable for use with Spanish-speaking children from different cultural and language dialect backgrounds. The test developed in the EL CIRCO project will be appropriate for children from Mexican, Puerto Rican, and Cuban backgrounds.
2. A variety of instruments to emphasize a differentiated view of human abilities. Hopefully, the collection of 13 instruments will discourage some of the misuse associated with single score measures.

3. Sufficient pilot testing to eliminate cultural or regional biases. Pilot testing groups have been selected to represent the three major cultural groups, as well as regional and SES differences.

4. The development of norms representative of the Spanish-speaking population of 4-, 5-, and 6-year old children in preschool programs in this country. Although many bilingual educators list this as a minimal criteria for selecting tests for use with Spanish-speaking children, to our knowledge representative norms have never before been attempted.

5. Major consideration of the construct of Spanish as a language and of syntax and lexical differences between Spanish and English. One fault of most translations is that these differences are not considered.

6. Adequate guards against test score misinterpretation and test misuse. Because the CIRCO tests in many ways resemble the CIRCUS tests, there is a tendency to interpret the scores on the CIRCO instruments relative to the performance of English-speaking students on the similar CIRCUS measure. By the technical definition the tests are not parallel and cannot be equated.

The cost for a development effort meeting these standards would be prohibitive if we did not have the materials and experience of CIRCUS as a base from which to work. The CIRCUS effort grew out of the concerns and interests of ETS staff members in providing valid and reliable measurement devices that can be used by teachers in planning and implementing their instructional programs. The circus theme that runs through the measures provides content that young children seem to respond to positively and to enjoy; in addition, the circus clowns, acrobats, and animals

are generally free from racial, geographic, socioeconomic, and sex bias.

CIRCUS is built upon a philosophy that measurement can and should be an integral part of early education and that the measurement needs of early education can best be met by an assessment battery emphasizing the complexity of the child and the broad range of skills, attitudes, and achievement that characterize his development. The CIRCUS series includes an array of instruments, each measuring an important dimension of the preschooler's capabilities. The results are presented in the form of a descriptive profile of each child's educational strengths and weaknesses. Thus, CIRCUS provides the early childhood educator with types of diagnostic and evaluative information not previously available. The CIRCUS instruments have been designed to be usable within the limitations of time, financial resources, and measurement expertise typical of many nursery schools and kindergartens without sacrificing the technical quality and validity demanded by decisions affecting the development of children.

Enough work has been completed on the CIRCUS battery in its English language form to consider parallel development of a Spanish language battery--CIRCO. Furthermore, the social conditions and measurement gaps which led ETS to develop the CIRCUS battery for English-speaking preschoolers are even more compelling in the case of children from Spanish-speaking backgrounds.

CIRCUS cannot serve one large segment of America's preschool population. Approximately 6 million persons (or 65 percent of the nation's 9.2 million persons of diverse Spanish origins) use Spanish at home. Among the school age children of Spanish origins (children 5 to 19 years old), 2.2 million (or 43 percent) speak only Spanish at home. There are more than 650,000 Spanish-speaking children between the

ages of 4 and 6, the target ages for CIRCUS (U. S. Bureau of the Census, 1973). Children whose first language is Spanish have often either been denied the benefits of diagnostic testing altogether or suffered from being misclassified and mislabeled on the basis of a test administered in an unfamiliar language. The extension of CIRCUS to include development of a Spanish version--CIRCO--can serve as a step toward more equal educational opportunity for a large and important group of young children.

In our planning for the development of a Spanish version of CIRCUS and later in a review of our project design by a blue ribbon panel of experts in the fields of early childhood education, research design, sampling and bilingual education, several philosophical and technical issues emerged. Not the least of these issues was the question of whether tests for Spanish-speaking preschoolers should be developed at all.

In recent years, a significant number of educators have voiced the opinion that misinterpretation and misuse of test scores is so prevalent that the risks of testing outweigh the advantages. National organizations such as NEA, APGA, and the Association of Black Psychologists have adopted resolutions calling for a moratorium on all or certain types of standardized testing (NEA, 1972; Samuda, 1975). Testing, in general, has some very bad connotations to much of the Spanish-speaking population of this country. These negative feelings are not without basis. Research reported by Carter (1970) indicates that Mexican-American children are overrepresented, in proportion to the general population, in EMR and other "Special Education Classes." Studies by Mercer (reported in Friedman, 1973) and by Palomares and Johnson (1966) support the opinion of most Spanish background Americans that, in fact, many of these students have

been placed in EMR classes on the basis of test scores which are not substantiated by other criteria.

The problem is summarized by Friedman (1973).

> "The vast majority of Chicano children bear the stigma of subnormal intelligence scores in their "cum-folders" all through the school years. Teachers who look at these scores for guidance have low expectations of the Chicano children, which often become self-fulfilling prophecies. The children fall farther and farther behind as they travel through the grades, frequently dropping out of high school. The Chicano dropout rate is one of the highest in this country. Thus, the educational system and testing helps to keep our people impoverished, unemployed, and ignorant.
>
> ...
>
> "To sum it up: The defenders of the current system of standardized testing use elitist arguments which are realistic-sounding only to the extent that the educational system is actually elitist in practice. 'The Chicano child is destined to fail in school. Therefore, it is only fair that he or she fail the tests also.' In fact, this proves that the tests are valid. And in turn, the tests prove that the schools are not to blame for failing the child. 'He or she just isn't bright enough.' This is the circular argument, corresponding to the vicious circle of low-expectation and low achievement in which the Chicano child is caught."
> (P. 280, 282)

Although the history of test misuse with minorities is tragic, most educators (including most Spanish-speaking educators) agree that the elimination of standardized testing is not an acceptable or feasible long range solution. As Henry Dyer (1961) put it:

> "Some of our brasher critics have argued that, since tests are so widely misused, they do constitute a menace to sound education and therefore should be abolished. This argument is specious. It is the same as saying that automobiles should be abolished because they are a menace to human life when reckless drivers are at the wheel." (in Samuda, 1975, p. 10.)

In response to the NEA resolution suggesting a moratorium on all standardized testing, William Coffman (1974) wrote a very eloquent rebuttal in which he stressed that while a moratorium may be a temporary solution, the best solution is for test developers to examine closely the reasons for test misuse and for the moratorium resolutions and to work toward new developments which will minimize test misuse. Two of the specific suggestions of Coffman are incorporated in the CIRCUS and CIRCO systems:

1. Interpretive scores which lend themselves to gross misinterpretation, such as mental age or grade equivalent, should be abandoned. (APA, 1973).
2. Experts (test publishers) should provide more direct interpretation as part of the score report. (Coffman, p. 5).

Many of the complaints against present tests (even when Spanish directions are used) stem from a) the fact that cultural differences are not considered, and b) norms groups are not representative of the Spanish-speaking population. These problems are extremely difficult and expensive to correct after a test is published, but are relatively easy to correct in test development. In the CIRCO development, an Advisory Committee including representatives of the three major Spanish-speaking

cultures in this country reviews all test directions and items for cultural and language relevance and often suggests alternative vocabulary or illustrations for new items. All instruments will be pilot tested, then field tested, and normed using samples selected to be representative of 4-, 5-, and 6-year old Spanish-speaking children in the United States.

I agree with Coffman (1974) that it is probably impossible to totally eliminate the misuse of tests, but with conscious attention to prior misuses and to the causes for those misuses, the test development process offers the most appropriate means of minimizing future misuse.

Method of Development

A second issue of major importance in considering an assessment system for Spanish-speaking preschoolers is the method of test development. With an existing test series developed for English-speaking children, the most obvious approach to a Spanish version is translation. In both professional opinion and research findings, one finds much conflict in assessing the value of tests developed in English and later translated into Spanish.

Zirkel (1975) points to several studies which indicate that Spanish translations produce higher scores for Spanish-speaking children than English versions of the same tests.

> "When Mahakian, for example, administered a standardized reading test in both original and translated versions to 210 Spanish-speaking children in grades 1 through 7, 83 percent achieved higher total scores in Spanish, with decreasing differences found in ascending grades. Thonis similarly found that sixteen out of nineteen Mexican-American students scored higher when tested

> with a Spanish translation of the Peabody Picture Vocabulary Test than they did with the standardized English version. According to Davis and Personke, the differences between Spanish and English administrations of the MRT were mostly nonsignificant for a group of Mexican-American 1st graders. However, the mean scores on the subtest that most appropriately reflected their language background (Word Meaning) revealed a significant difference favoring the Spanish version." (P. 295-296).

Most bilingual educators, however, agree that translation alone is not an adequate solution to testing Spanish-speaking children.

> "...it is no solution to simply translate standardized tests or instructional materials into Spanish." (Mendoza-Friedman, 1973).
>
> "The development of tests appropriate to Spanish-speaking children is far, far more than simply translating existing tests." (Finch, 1971).
>
> "The so-called translation of test items into a non-English language...indicates an urgency to develop appropriate instruments in the mother tongue." (Bernal, 1975).

Some of the weaknesses of translations frequently cited are

1. translation does not eliminate items which are culturally biased,
2. translations usually do not allow for local dialect variations, and
3. translations often affect item difficulty due to differences of lexical items across languages.

These limitations should not lead one to conclude that there should be no relationship between measures for Spanish-speaking children and

measures for English-speaking children. Quite the contrary. There are many research and diagnostic benefits to be gained through the development of English and Spanish language versions that are parallel in the constructs they measure. The task, then, is to translate and adapt the tests in such a way that the construct measured is retained without the introduction of extraneous difficulties due to language or cultural differences.

One of our early activities in the CIRCO project was to have selected instruments translated by three independent teams representing Mexican-American, Puerto Rican, and Cuban dialects. These initial translations illustrated several dialect and cultural differences requiring test adaptation. With the help of our Cultural Advisory Committee, several different methods of dealing with cultural and language differences are being tried. Perhaps some examples will illustrate the problems and (hopefully) the solutions.

Example 1

Problem: In the test "How Much and How Many" an item asks which picture has the fewest ice cream cones. Our translators could not agree on a phrase for ice cream cone that would be understood by all three culture groups.

Solution: Ice cream cone is not essential to the construct being measured (fewest), therefore, we may substitute an object with a name common to all three cultures (apples or dogs).

Example 2:

Problem: Three different Spanish phrases are used for "Turn the page."

Solution: Print all three and allow the test administrator to choose the phrase most appropriate for the children being tested.

Example 3

Problem: The Spanish translation of "Please," Por favor, is only used when requesting a favor. It should not be used in giving instructions or commands.

Solution: Whereas "Please turn the page" is appropriate in English, an appropriate Spanish instruction would be "Pasen la pagina," not "Pasen la pagina, por favor."

Example 4

Problem: In "Copy What You See," a measure of perceptual motor coordination, children are asked to copy the letter "W." The letter "W" occurs very infrequently in Spanish.

Solution: Substitute the letter "M" which requires similar motor skills to copy and occurs with more equal frequency in English and Spanish.

While I think you will agree that, in the examples, neither the problems nor the solutions are very dramatic, they are illustrative of the types of biases that creep into a test when there is not careful attention to cultural and dialect differences.

It should perhaps be noted that while translation with cultural adaptation of items is an appropriate development method for some measures, not all measures can be developed in this way. In the CIRCUS series is a measure of phoneme discrimination. Items are of the type:

"Here is a man, a pan, and a can. Mark the picture of the man."

When translated, the words for man, pan, and can no longer differ only in the initial consonant. The Spanish version of this measure will, therefore, have all new items, but the construct measured, and probably the item type, will be the same. The CIRCUS series includes a measure of functional language. The Spanish version of this test will also be a new development but for a different reason: The most important syntax characteristics of English and Spanish are different. For example, every noun in Spanish is either masculine or feminine and the person learning Spanish must learn the proper article and pronoun forms for gendered nouns. This concept is, of course, not measured in the English version of the functional language instrument.

Perhaps these examples illustrate the need for flexibility and a variety of approaches to test development. Some tests may be adapted for use with Spanish-speaking children by a careful translation of the test directions. Other instruments must be considered a new test development borrowing concepts and approaches from the English version only where equally appropriate to both populations.

Score Interpretation

The CIRCUS series offers a distinctive score interpretation

service. For each measure, a child will receive one of four- or five-sentence interpretations. These sentence reports might range from "very competent in recognizing letters and numbers" to a report for chance level scores that says, in effect, "we don't think this was a very reliable or valid measure of this child's ability." The sentence reports were developed by ETS based upon a logical analysis of the content of each test and upon the performance of preschool and kindergarten children in a national norming administration. The CIRCUS test scores, then, have elements of both a content-referenced and a norm-referenced interpretation.

When a Spanish version of CIRCUS is considered, score interpretation becomes a problem.

One of the usual weaknesses in testing minority students has been the lack of appropriate norms. Fishman (1964) in his "Guidelines for testing minority group children" suggests the development of minority group norms as a basic requirement of all tests used with minority children.

The CIRCO project is committed to a development of national norms for the ten instruments being developed. These norms will allow a comparison of a child's status among other Spanish-speaking children in this country of the same age and economic status. It is the interpretation of the Spanish-speaking child's scores with respect to the scores of English-speaking children that gives the psychometrician gray hairs.

The initial temptation is simply to take the CIRCO score, enter the CIRCUS norms table, and say that this Spanish-speaking child scores at the ___ percentile of English-speaking children who took the same test.

While this approach is logical, it is based upon some very questionable assumptions. The first assumption is that a raw score of 34 on the Spanish version of a test represents the same quantity of a concept as a raw score of 34 on the English version. It is probably impossible to demonstrate that this assumption is true. Our usual method for handling this situation is to equate the two tests, converting scores to a common scale. The scaled scores will have the property of equal measure of units regardless of the test form from which the score is derived. Unfortunately, the CIRCO and CIRCUS tests do not meet the technical criteria for being parallel and cannot, therefore, be equated.

> "...two tests may be considered parallel if, after conversion to the same scale, their means, standard deviations, and correlations with any and all outside criteria are equal." (Gulliksen, 1975). (underline added)

Obviously, the correlations of the two test forms with language will not be the same. With nonparallel tests, comparable scores are possible, but most methods require that both tests be given to the same population (see Angoff, 1971). It is doubtful that a population for which CIRCO and CIRCUS would be equally appropriate can be identified.

Given the present state of the art, we must conclude that any equating of CIRCO and CIRCUS test scores would be of questionable validity and we must caution against and discourage the logical comparison.

Even if we give up on direct score comparisons, we still have a problem in the sentence report interpretations for CIRCO. Perhaps an example will illustrate:

> On a certain CIRCUS test, students who score two or more standard deviations above the mean of the norms group are given a sentence report,

> "Exceptionally competent..." Students who score between 0 and 2 standard deviations receive a report, "Generally competent..." Let us assume, as is generally the case, that the Spanish-speaking population scores are lower by about one standard deviation. For the Spanish student, who is two standard deviations above the CIRCO mean but only one standard deviation above the CIRCUS mean, do we give a sentence report of "Exceptionally competent" or "Generally competent."

In light of our previous caution against a comparison of CIRCO scores with English norms, the tendency would be to interpret in terms of the Spanish population and give a sentence report saying "Exceptionally competent." Some Hispanic educators, however, believe that the type of double standard will only perpetuate the difference. What is "generally competent" for an English-speaking child cannot be "exceptionally competent" for the Spanish-speaking child.

The solution to this dilemma is not easy. To void the double standard, the sentence reports will have to be linked to some criteria other than the normative comparisons. For some constructs, a Gagné-type hierarchy analysis may suggest criterion levels. In other areas, criterion levels may be set by preschool and kindergarten teachers based upon their experiences. Some current research at ETS is investigating the appropriateness of this approach.

This paper has presented some of the issues of philosophy, test development, and score interpretation emerging from an attempt to adapt a test series, originally developed in English, for Spanish-speaking children.

The approaches to these issues mentioned here do not represent solutions, but perhaps do represent steps toward more equitable testing for young Spanish-speaking children.

References

American Psychological Association. Standards for the development and use of educational and psychological tests. APA Monitor, September, 1973.

Angoff, W. H. Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational Measurement. Washington, D. C.: American Council on Education, 1971.

Bernal, E. M., Jr. A response to "Educational uses of tests with disadvantaged subjects." American Psychologist, January 1975, 93-95.

Carter, T. P. Mexican Americans in school: A history of neglect. New York: College Entrance Examination Board, 1970.

Coffman, W. E. A moratorium? What kind? NCME Measurement in Education, Spring, 1974, 5 (2), 1-7.

Finch F. L. Vamos: To develop a bilingual examination. Paper presented at the fifth annual TESOL convention, New Orleans, March 6, 1971.

Fishman, J. A., et. al. Guidelines for testing minority children. Journal of Social Issues Supplement, 1964, 20, 129-145.

Gulliksen, H. Theory of mental tests. New York: Wiley and Sons, 1950.

Mendoza-Friedman, M. Spanish bilingual students and intelligence testing. Education of the Spanish-speaking urban child: A book of readings. Springfield, Illinois: Charles C. Thomas, Publisher, 1975.

National Education Association. Addresses and proceedings of the one-hundred-and-tenth annual meeting held at Atlantic City, New Jersey, June 24-June 30, 1972, 110. Washington, D. C.: National Education Association, 1972.

Palomores, U. H., & Johnson, L. C. Evaluation of Mexican-American pupils for EMR classes. California Education, 1966, 3, 27-29.

Samuda, R. J. Psychological testing of American minorities: Issues and consequences. New York: Dodd, Mead & Company, 1975.

Tests and Measures in early childhood: Report I (ETS PR 72-17). Princeton, N. J.: Educational Testing Service, 1972.

U. S. Bureau of the Census. Census of the population: 1970, subject report persons of Spanish original (Series PC [2] -1C).

Zirkel, P. A. Spanish-speaking students and standardized tests. Education of the Spanish-speaking urban child: A book of readings. Springfield, Illinois: Charles C. Thomas, Publisher, 1975.